Challenge: Others … … …

Hao Dong

Peking University

Challenge: Others

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

InGAN

- InGAN: Capturing and Remapping the "**DNA**" of **a** Natural Image

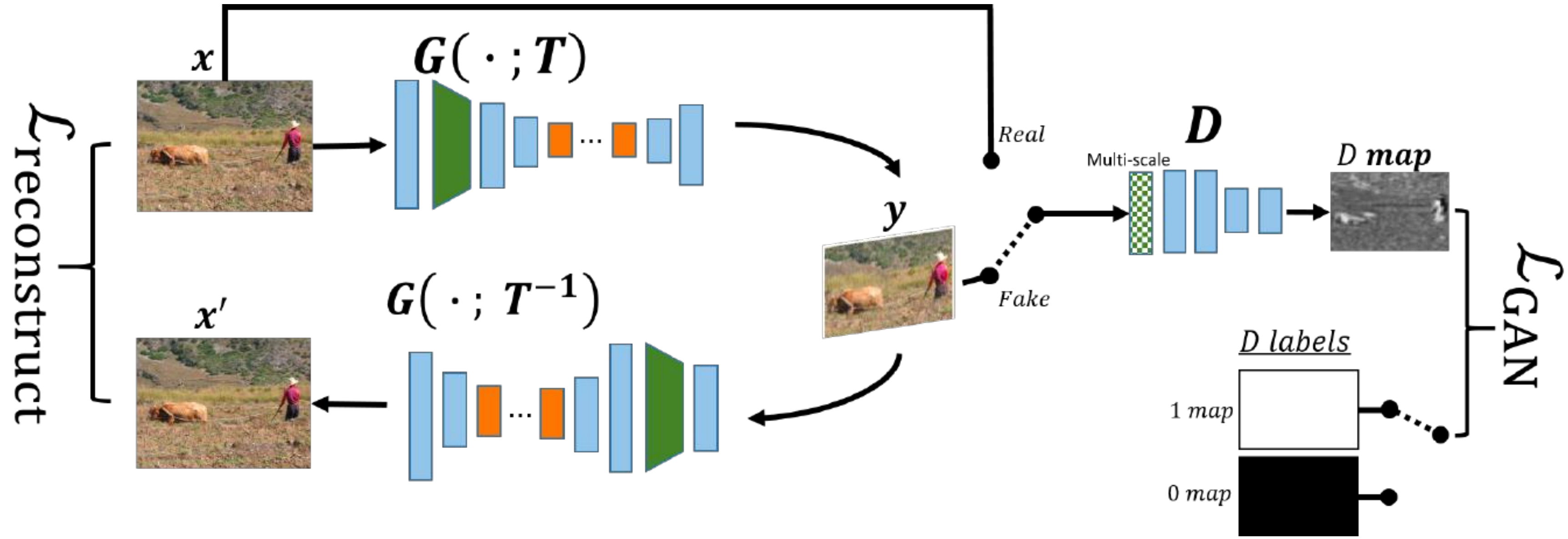
Conditional generative model

InGAN

- Architecture

$$\mathcal{L}_{\texttt{InGAN}} = \mathcal{L}_{\texttt{GAN}} + \lambda \cdot \mathcal{L}_{\texttt{reconst}}$$

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{y \sim p_{\text{data}}(x)}[(D(x){-}1)^2] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[D(G(x))^2]$$

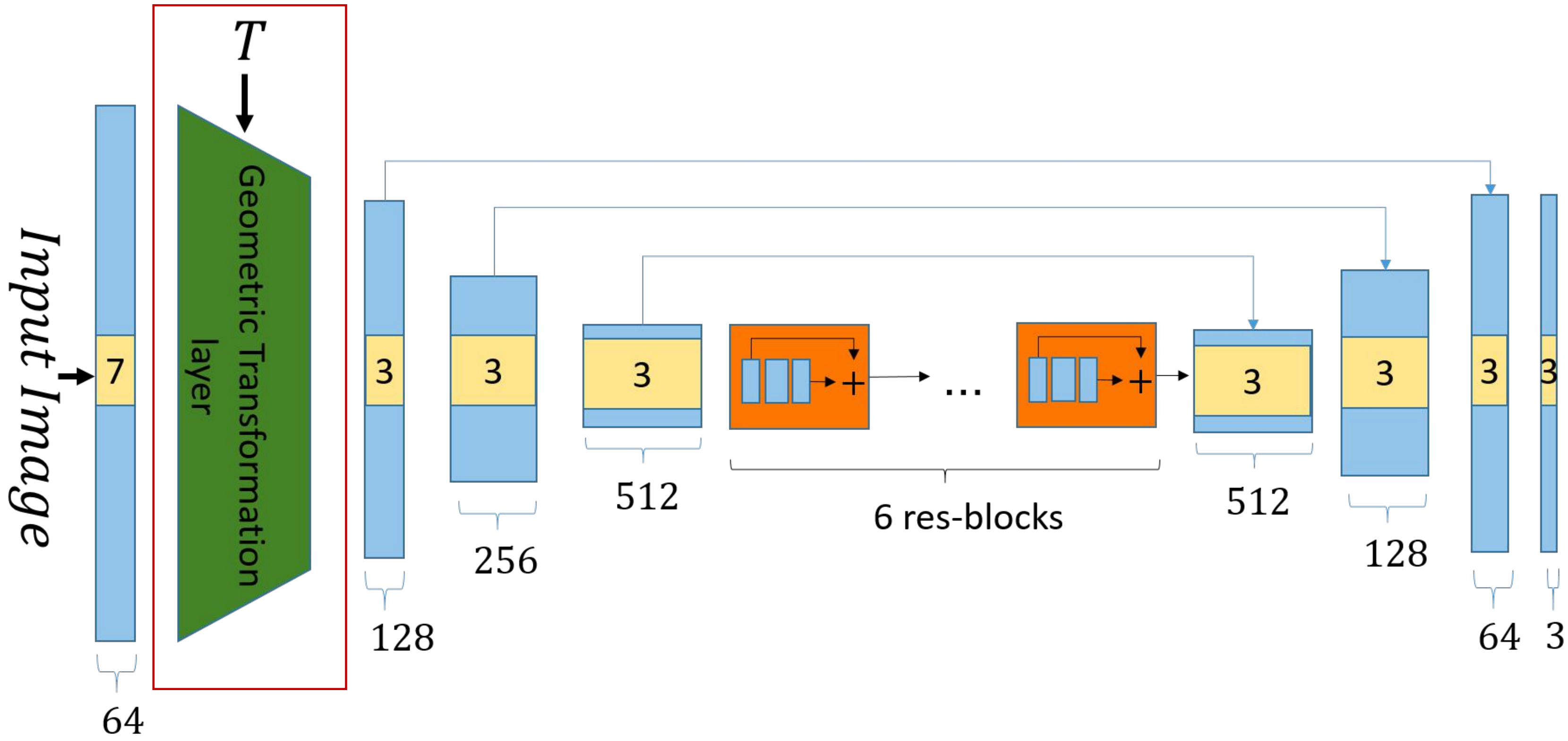


$$\mathcal{L}_{\texttt{reconst}} = \left\| G\left(G\left(x;T\right);T^{-1}\right) - x \right\|_1$$

InGAN

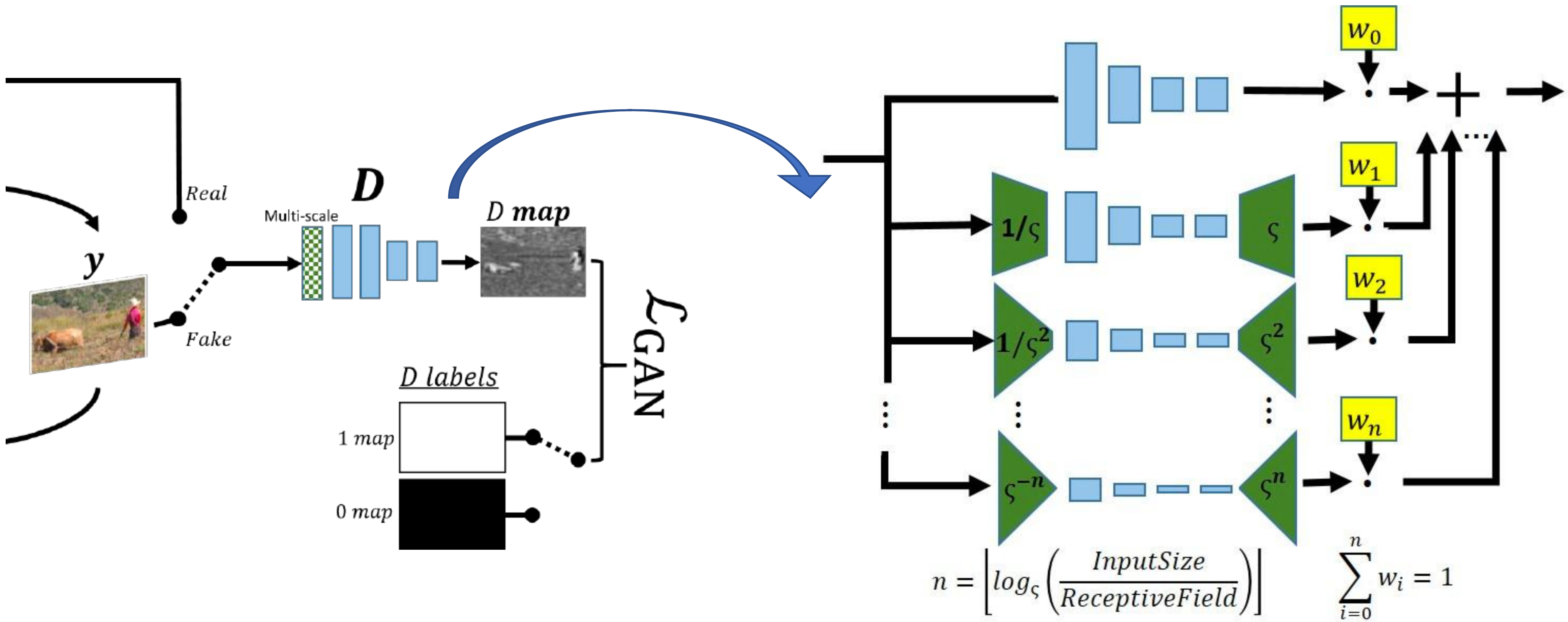
PEKING UNIVERSITY

- Generator architecture



InGAN

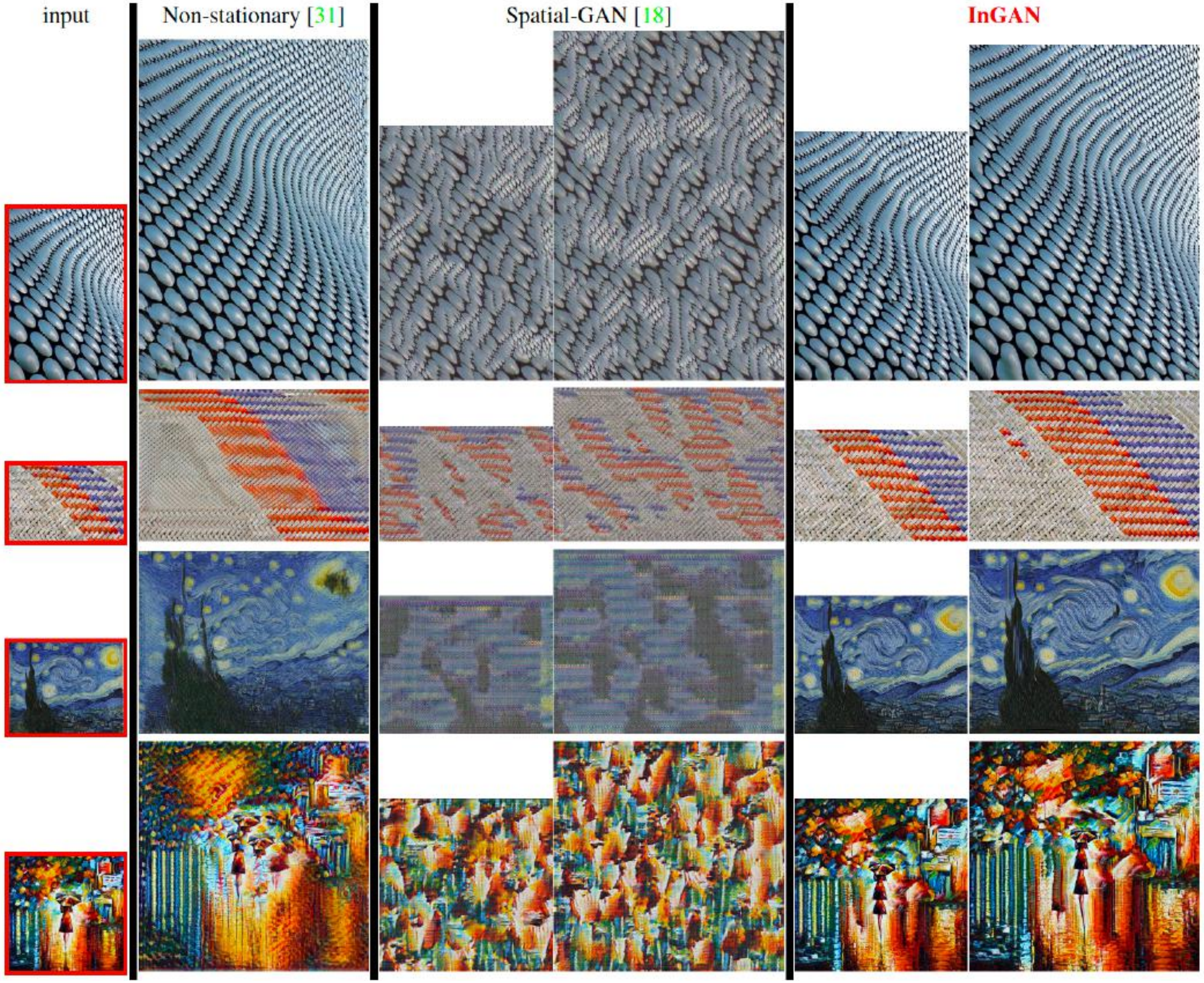
- Adaptive Multi-Scale Patch Discriminator



InGAN

Multiple Tasks:

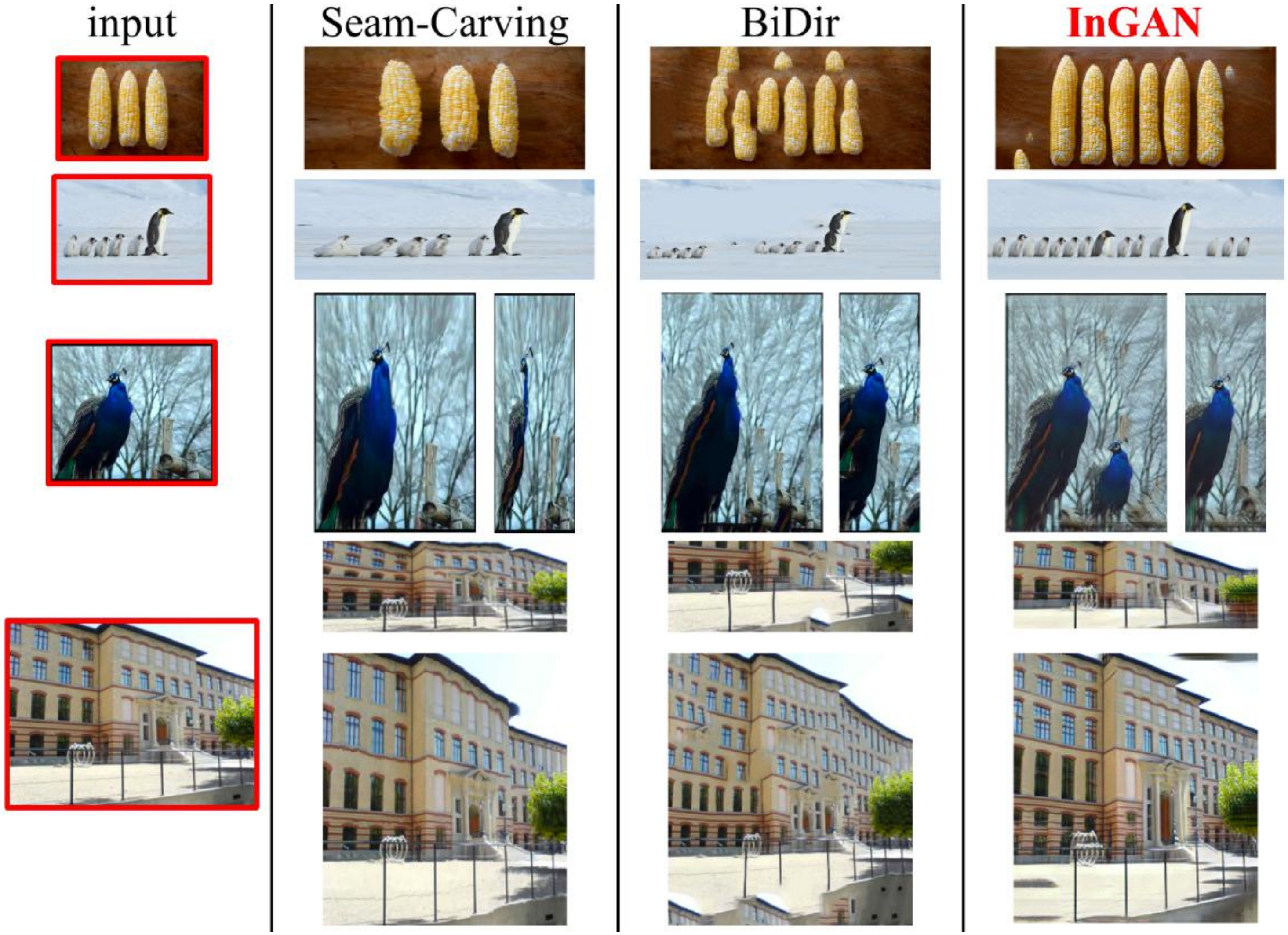
- Texture synthesis



InGAN

Multiple Tasks:

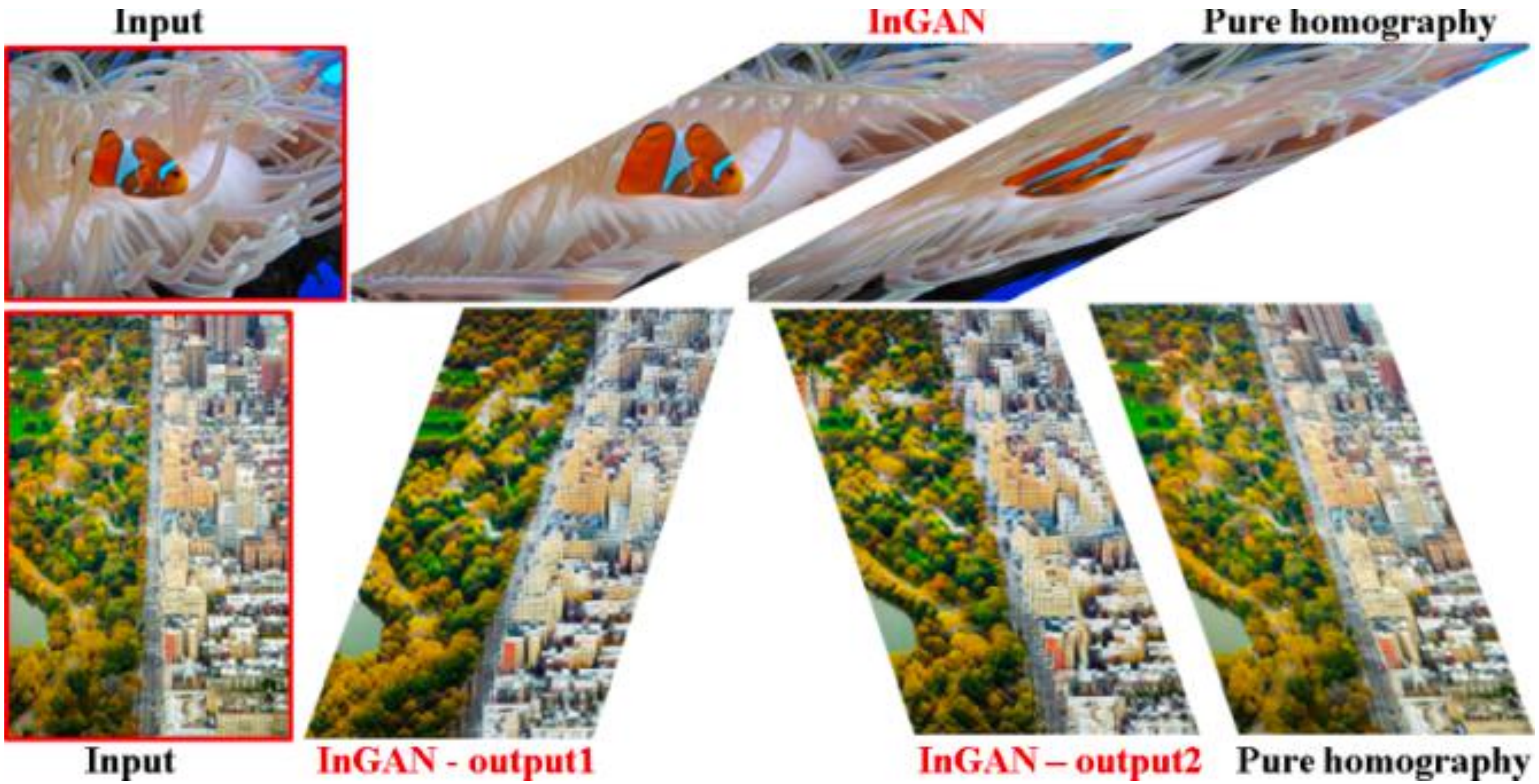
- Natural image retargeting



InGAN

Multiple Tasks:

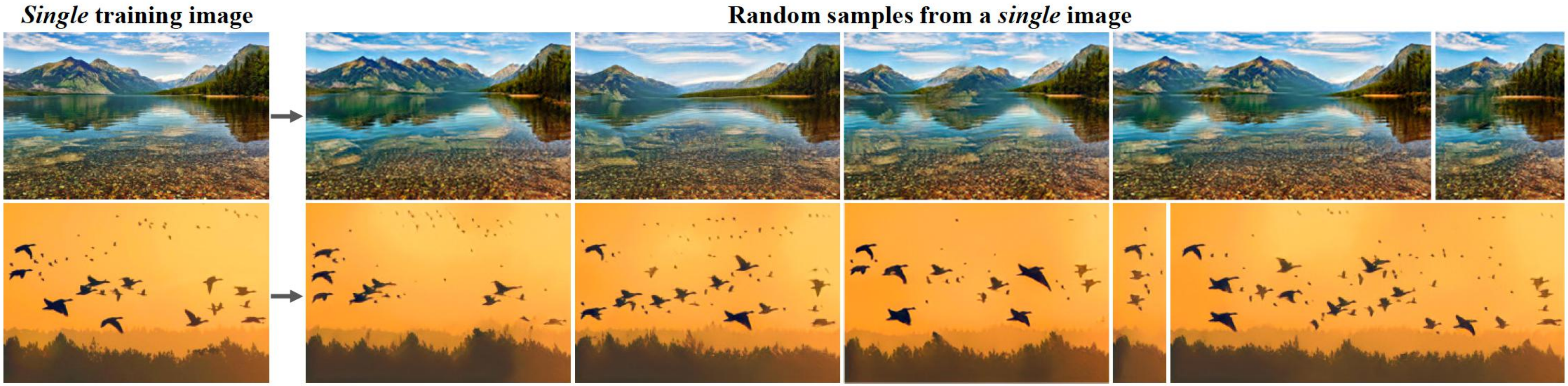
- Retargeting to Non-Rectangular Outputs



- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

SinGAN

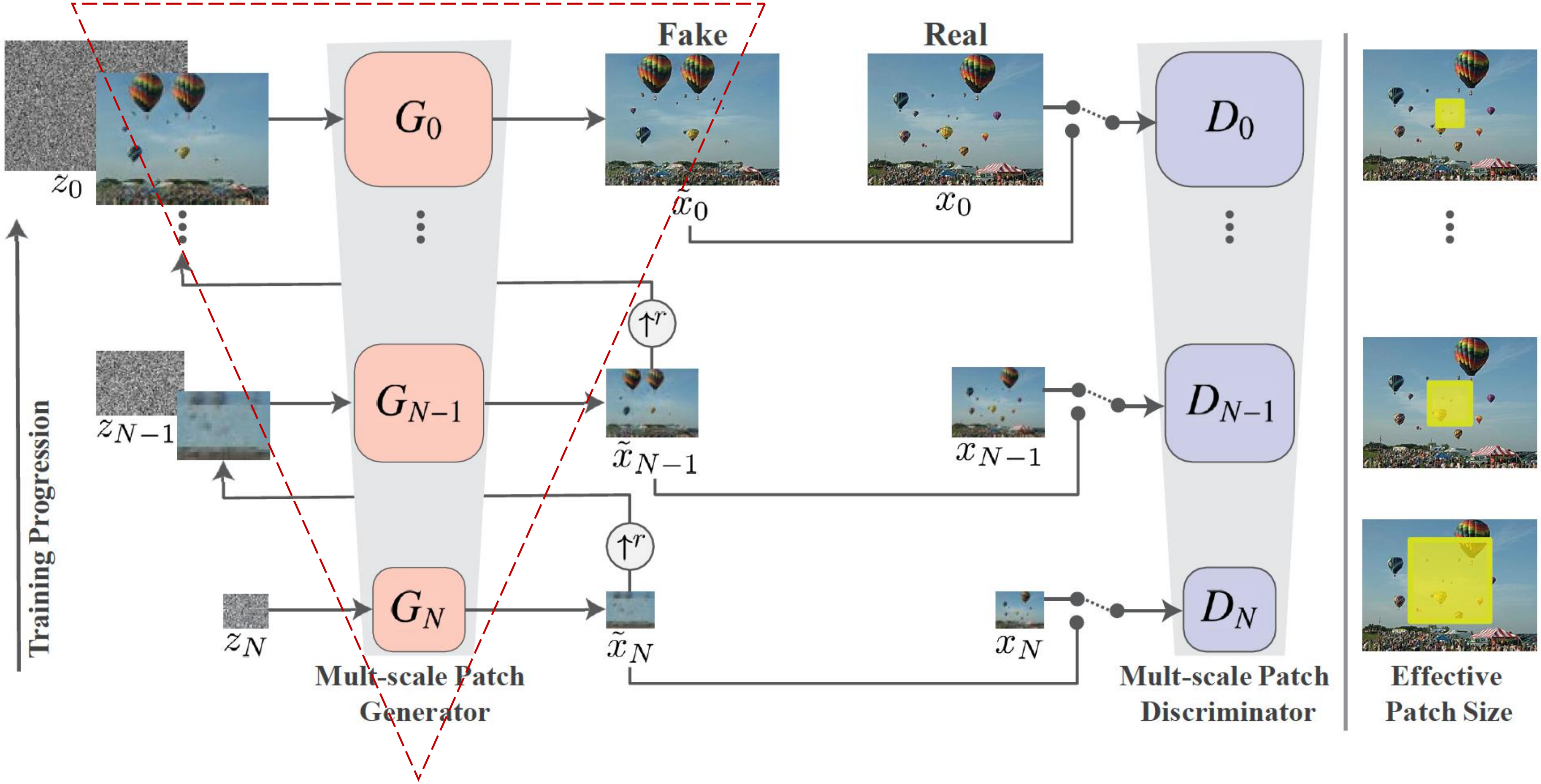
- SinGAN：Learning a Generative Model from a Single Natural Image



SinGAN: Unconditional VS. InGAN: Conditional

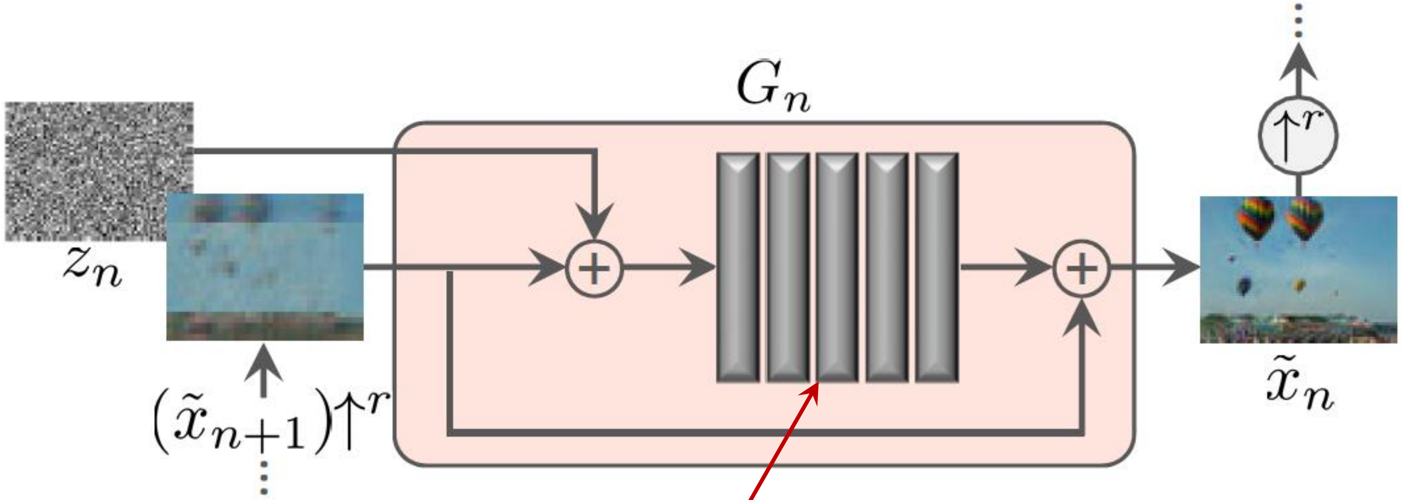
SinGAN

- SinGAN's multi-scale pipeline: A pyramid of GANs



SinGAN

- Single scale generation



$$\tilde{x}_n = G_n\left(z_n, (\tilde{x}_{n+1})\uparrow^r\right)$$

$$\tilde{x}_n = (\tilde{x}_{n+1})\uparrow^r + \boxed{\psi_n}\left(z_n + (\tilde{x}_{n+1})\uparrow^r\right)$$

SinGAN

- Training

  Sequentially train from the **coarsest** scale to the **finest** one

  Once each GAN is trained, it is kept fixed

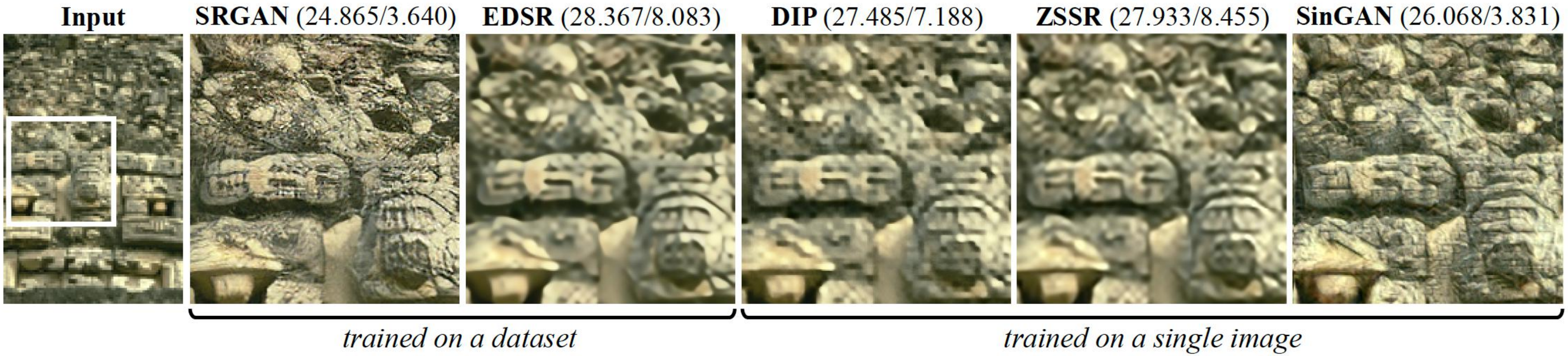
$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n)$$

WGAN-GP loss

$$\mathcal{L}_{\text{rec}} = \| G_n(0, (\tilde{x}^{\text{rec}}_{n+1}) \uparrow^r) - x_n \|^2$$

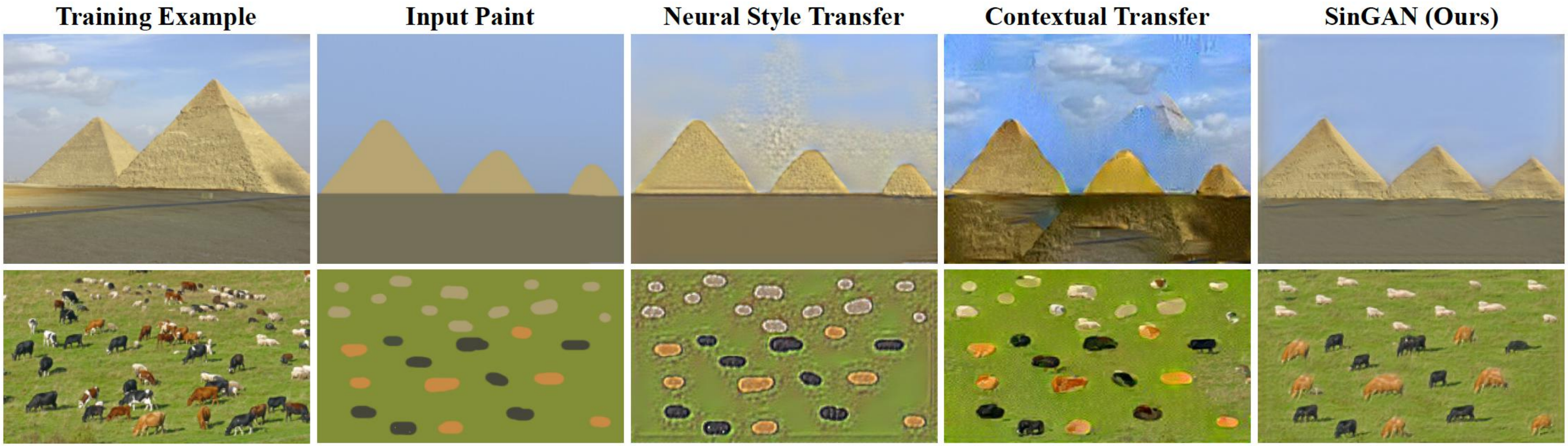
SinGAN

- Applications: Super Resolution



SinGAN

- Applications: Paint-to-Image



SinGAN

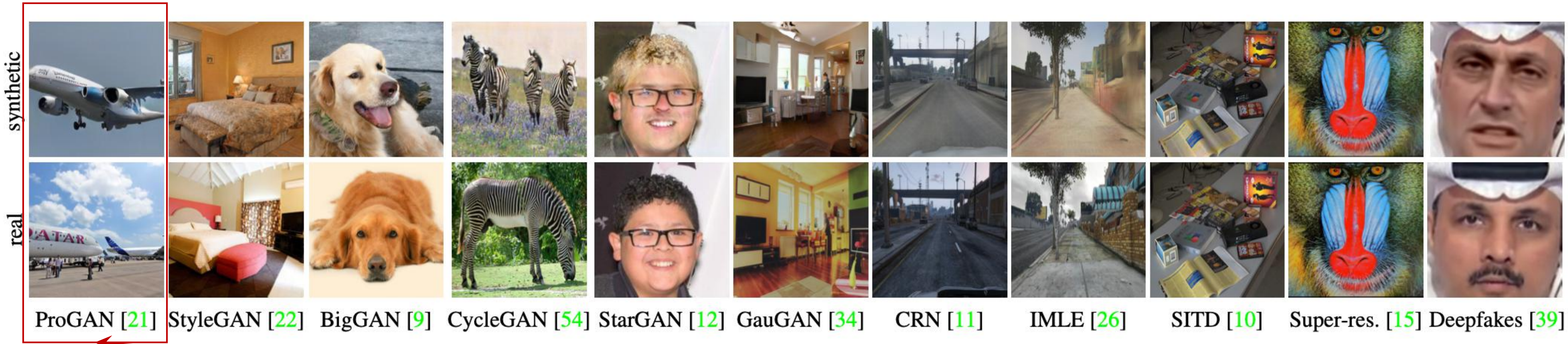
- Applications: Harmonization

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

What it is in the frequency domain

- CNN-generated images are surprisingly easy to spot... for now

Are CNN-generated images hard to distinguish from real images?



A classifier trained to detect images **generated by only one CNN** (ProGAN, far left) **can detect those generated by many other models** (remaining columns)

What it is in the frequency domain

- CNN-generated images are surprisingly easy to spot... for now

ForenSynth : A dataset of CNN-based generation models

Family	Method	Image Source	# Images
Unconditional GAN	ProGAN [21]	LSUN	8.0k
	StyleGAN [22]	LSUN	12.0k
	BigGAN [9]	ImageNet	4.0k
Conditional GAN	CycleGAN [54]	Style/object transfer	2.6k
	StarGAN [12]	CelebA	4.0k
	GauGAN [34]	COCO	10.0k
Perceptual loss	CRN [11]	GTA	12.8k
	IMLE [26]	GTA	12.8k
Low-level vision	SITD [10]	Raw camera	360
	SAN [15]	Standard SR benchmark	440
Deepfake	FaceForensics++ [39]	Videos of faces	5.4k

What it is in the frequency domain

- CNN-generated images are surprisingly easy to spot... for now

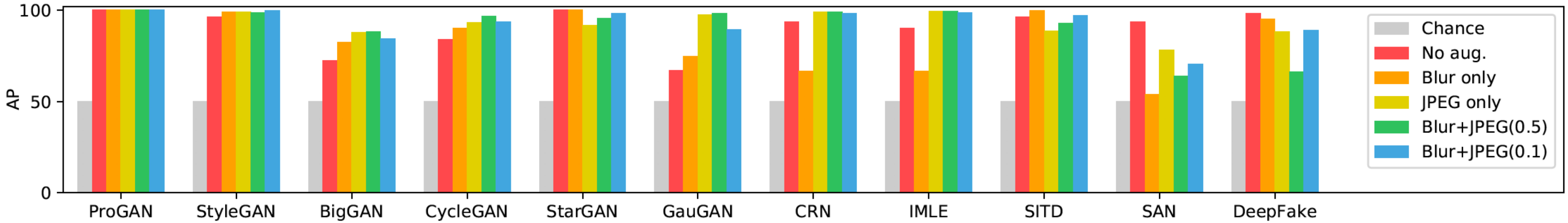
Effect of data augmentation

Family	Name	Training settings					Individual test generators											Total
		Train	Input	No. Class	Augments Blur	Augments JPEG	Pro-GAN	Style-GAN	Big-GAN	Cycle-GAN	Star-GAN	Gau-GAN	CRN	IMLE	SITD	SAN	Deep-Fake	**mAP**
Zhang et al. [50]	Cyc-Im	CycleGAN	RGB	–			84.3	65.7	55.1	100.	99.2	79.9	74.5	90.6	67.8	82.9	53.2	77.6
	Cyc-Spec	CycleGAN	Spec	–			51.4	52.7	79.6	100.	**100.**	70.8	64.7	71.3	92.2	78.5	44.5	73.2
	Auto-Im	AutoGAN	RGB	–			73.8	60.1	46.1	99.9	**100.**	49.0	82.5	71.0	80.1	86.7	80.8	75.5
	Auto-Spec	AutoGAN	Spec	–			75.6	68.6	84.9	**100.**	**100.**	61.0	80.8	75.3	89.9	66.1	39.0	76.5
Ours	2-class	ProGAN	RGB	2	✓	✓	98.8	78.3	66.4	88.7	87.3	87.4	94.0	97.3	85.2	52.9	58.1	81.3
	4-class	ProGAN	RGB	4	✓	✓	99.8	87.0	74.0	93.2	92.3	94.1	95.8	97.5	87.8	58.5	59.6	85.4
	8-class	ProGAN	RGB	8	✓	✓	99.9	94.2	78.9	94.3	91.9	95.4	98.9	99.4	91.2	58.6	63.8	87.9
	16-class	ProGAN	RGB	16	✓	✓	100.	98.2	87.7	96.4	95.5	**98.1**	**99.0**	**99.7**	95.3	63.1	71.9	91.4
	No aug	ProGAN	RGB	20			100.	96.3	72.2	84.0	**100.**	67.0	93.5	90.3	96.2	**93.6**	**98.2**	90.1
	Blur only	ProGAN	RGB	20	✓		100.	99.0	82.5	90.1	**100.**	74.7	66.6	66.7	**99.6**	53.7	95.1	84.4
	JPEG only	ProGAN	RGB	20		✓	100.	99.0	87.8	93.2	91.8	97.5	**99.0**	99.5	88.7	78.1	88.1	**93.0**
	Blur+JPEG (0.5)	ProGAN	RGB	20	✓	✓	100.	98.5	**88.2**	96.8	95.4	**98.1**	98.9	99.5	92.7	63.9	66.3	90.8
	Blur+JPEG (0.1)	ProGAN	RGB	20	†	†	100.	**99.6**	84.5	93.5	98.2	89.5	98.2	98.4	97.2	70.5	89.0	92.6

What it is in the frequency domain

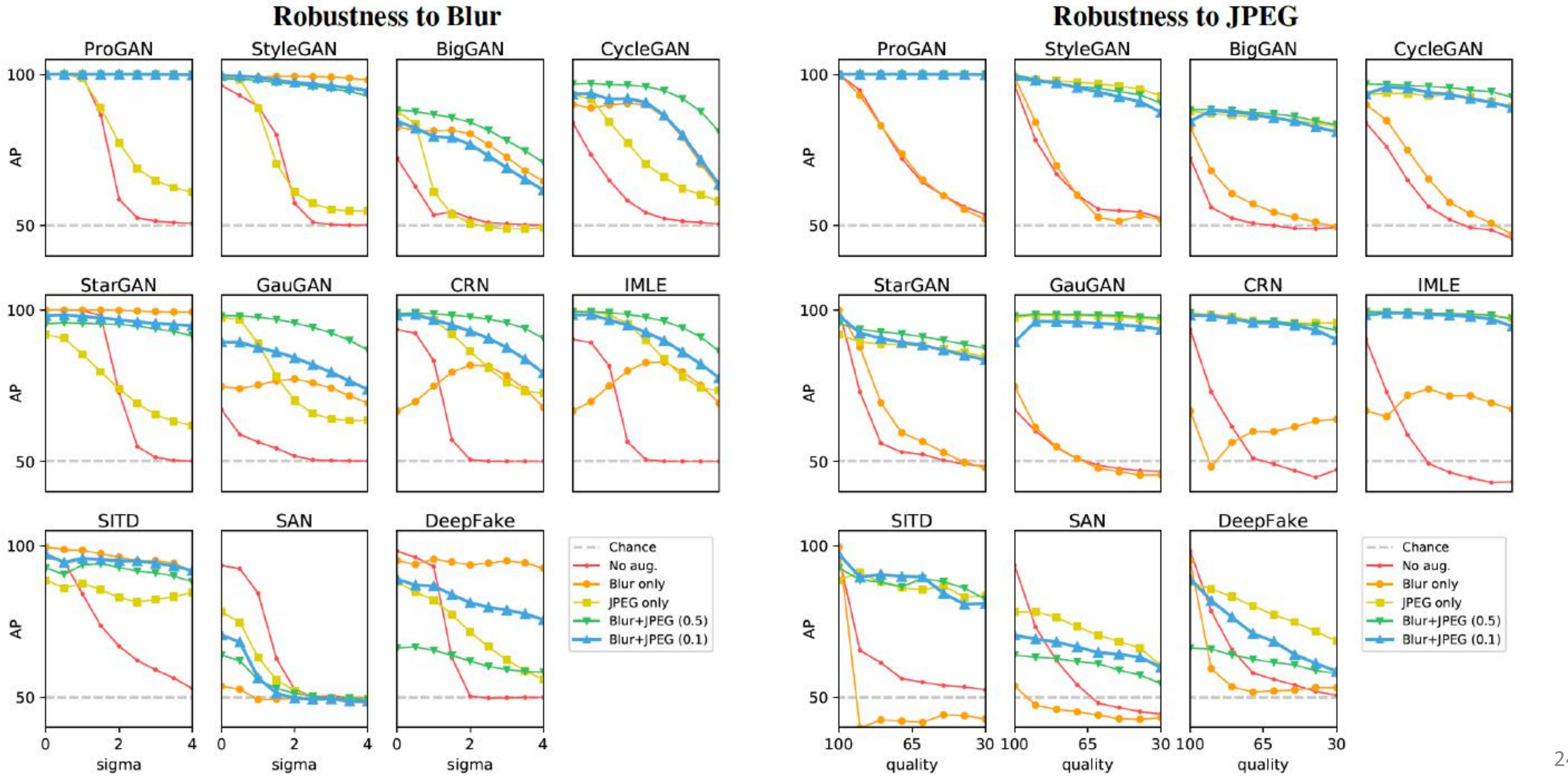
- CNN-generated images are surprisingly easy to spot... for now

Effect of data augmentation



What it is in the frequency domain

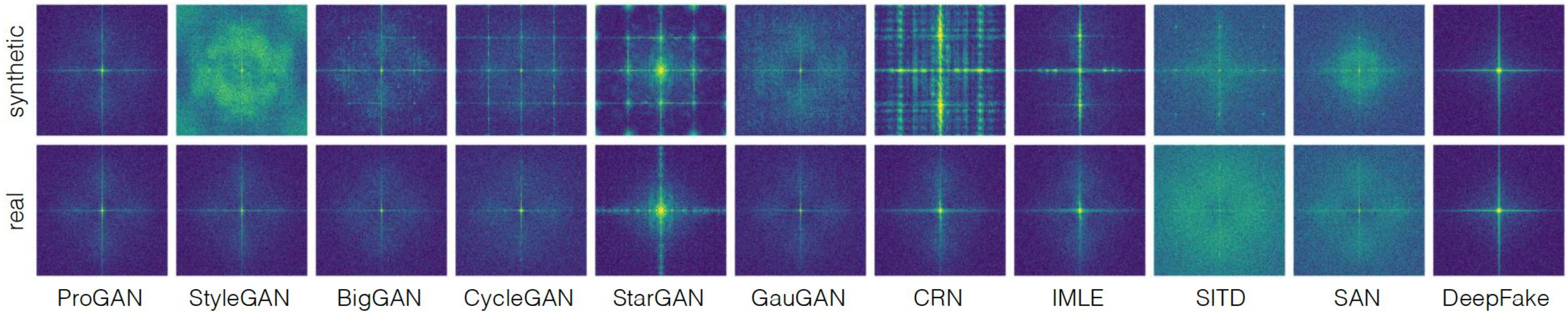
- CNN-generated images are surprisingly easy to spot... for now



What it is in the frequency domain

- CNN-generated images are surprisingly easy to spot... for now

  Frequency analysis on each dataset

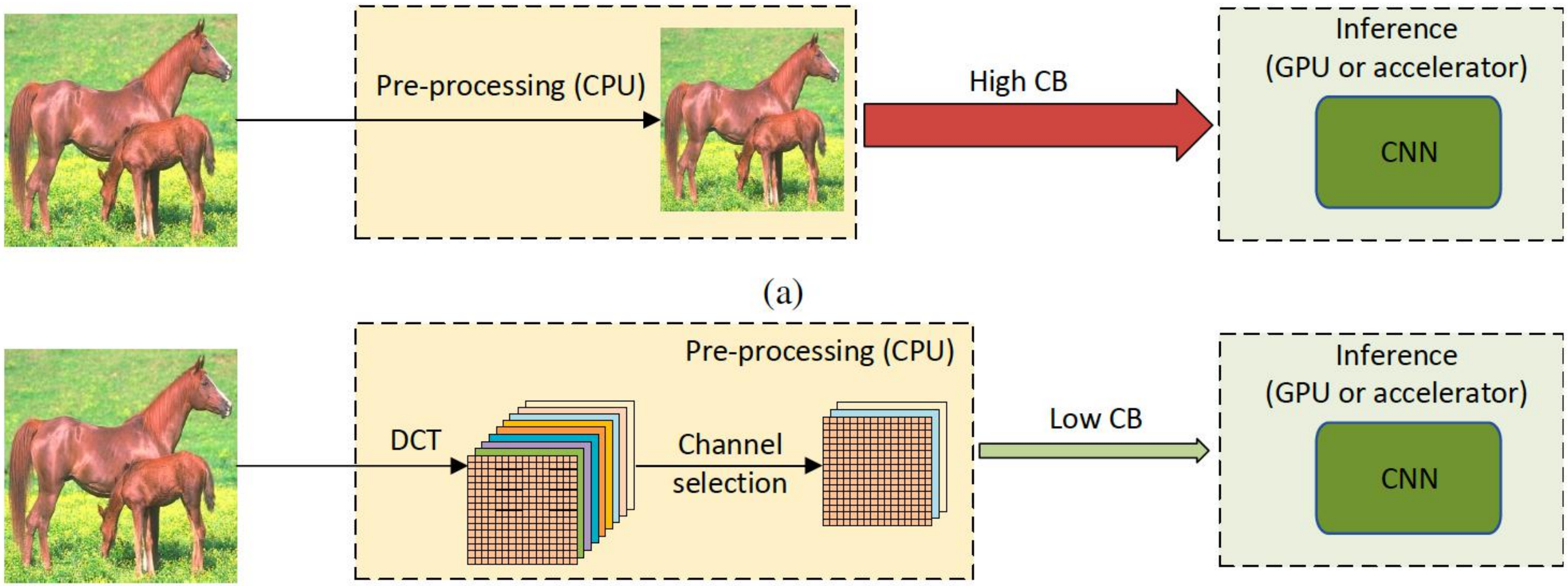


- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

What it is in the frequency domain

- Learning in the Frequency Domain

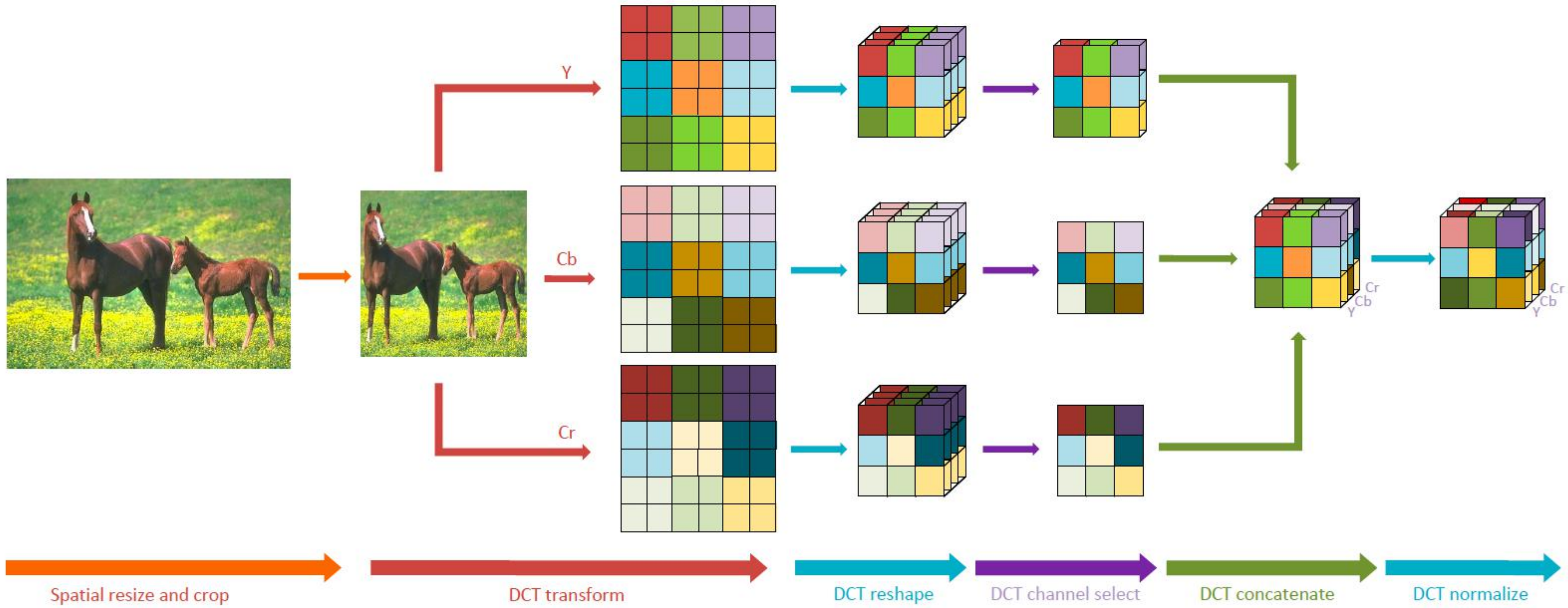
  Why in the frequency domain?



What it is in the frequency domain

- Learning in the Frequency Domain

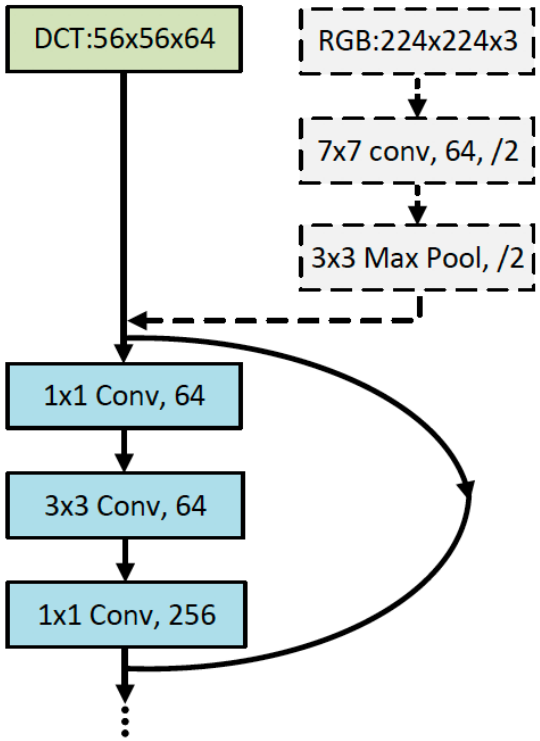
  Data pre-processing pipeline



What it is in the frequency domain

- Learning in the Frequency Domain

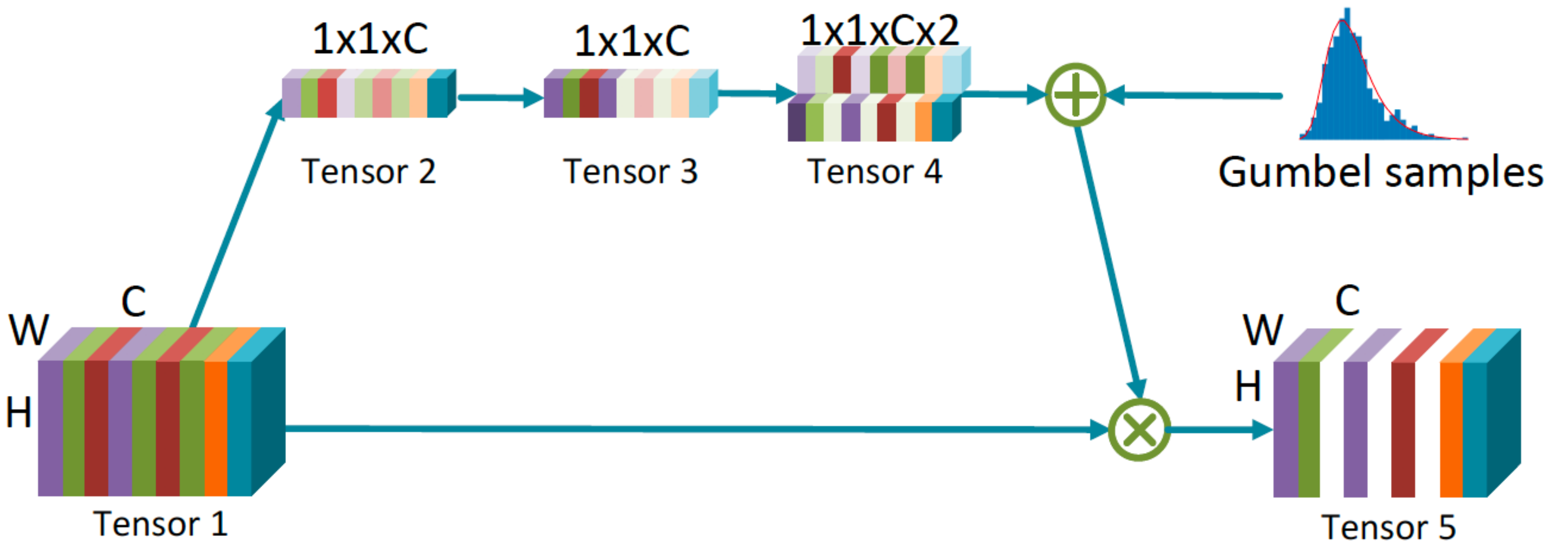
  How to convert into the frequency domain?



What it is in the frequency domain

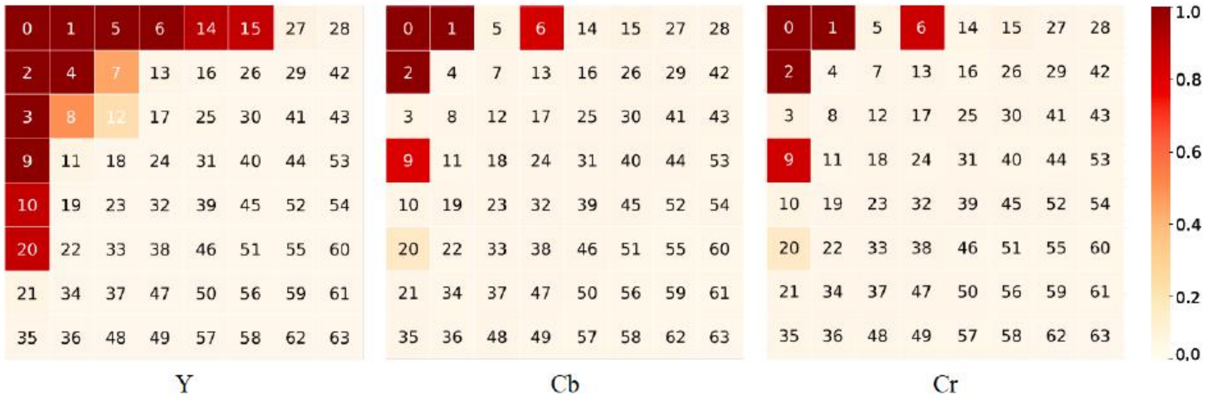
- Learning in the Frequency Domain

  Channel Selection

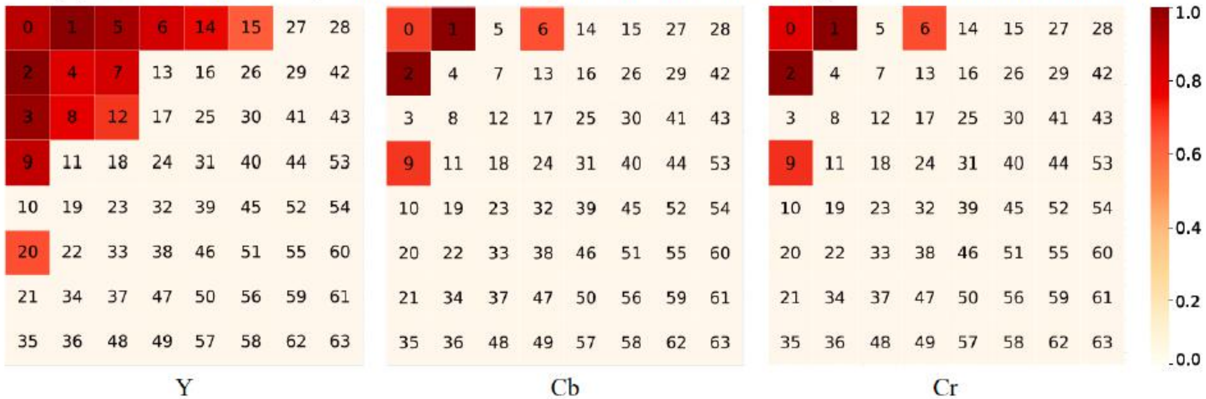


What it is in the frequency domain

- Learning in the Frequency Domain



(a) Heat maps of Y, Cb, and Cr components on the ImageNet validation dataset.



(b) Heat maps of Y, Cb, and Cr components on the COCO validation dataset

What it is in the frequency domain

- Learning in the Frequency Domain

Examples of instance segmentation results on the COCO dataset

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

What it learns

- GAN Dissection

How to visualise GANs?

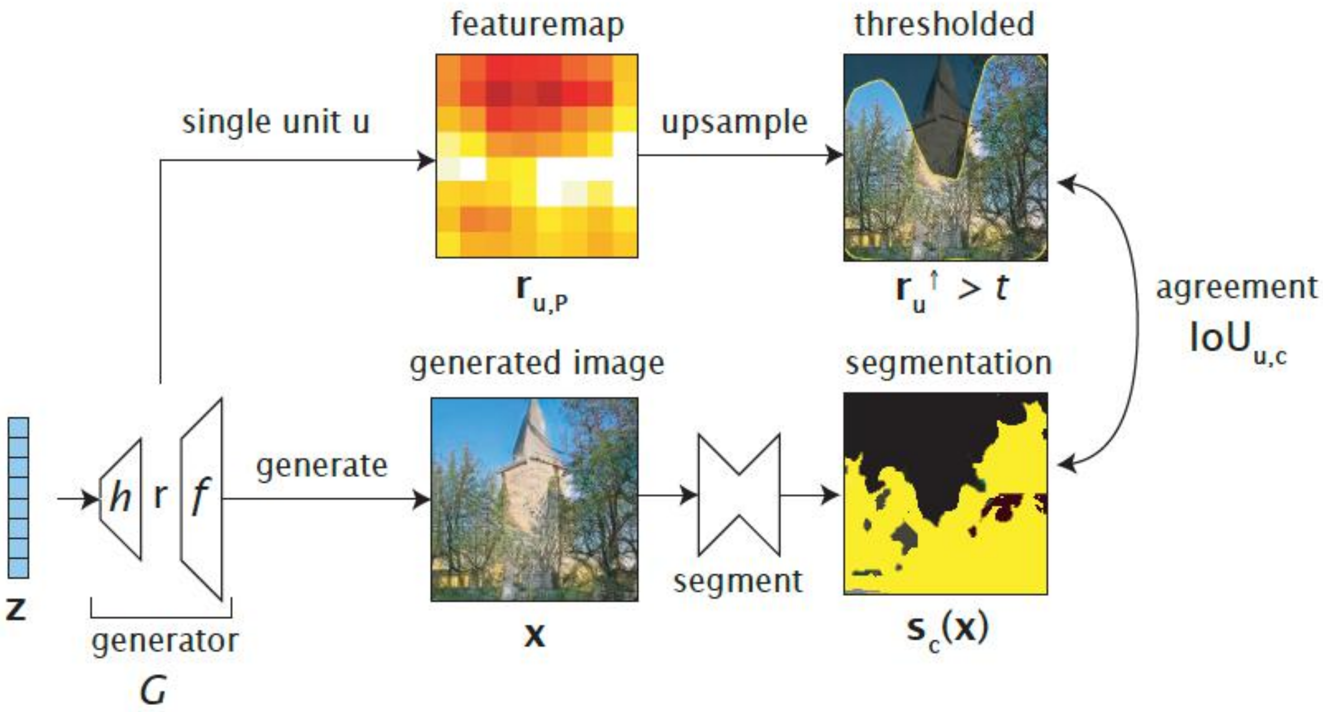
How to understand GANs?

Bau, David, et al. "Gan dissection: Visualizing and understanding generative adversarial networks." arXiv preprint arXiv:1811.10597 (2018).

What it learns

- GAN Dissection

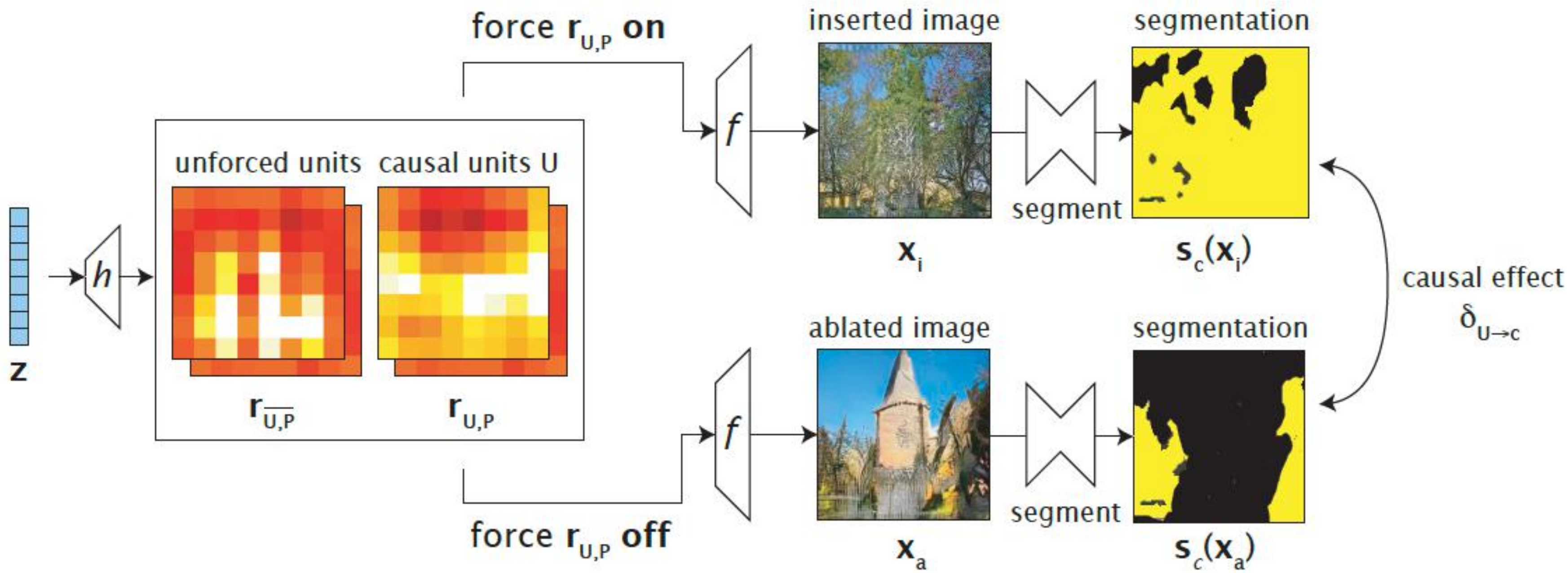
Analytical Framework: Characterising Units by **Dissection**



What it learns

- GAN Dissection

  Analytical Framework: Measuring Causal Relationships Using **Intervention**



What it learns

- GAN Dissection

  Finding concepts

(a) Generate images of churches

(b) Identify GAN units that match trees

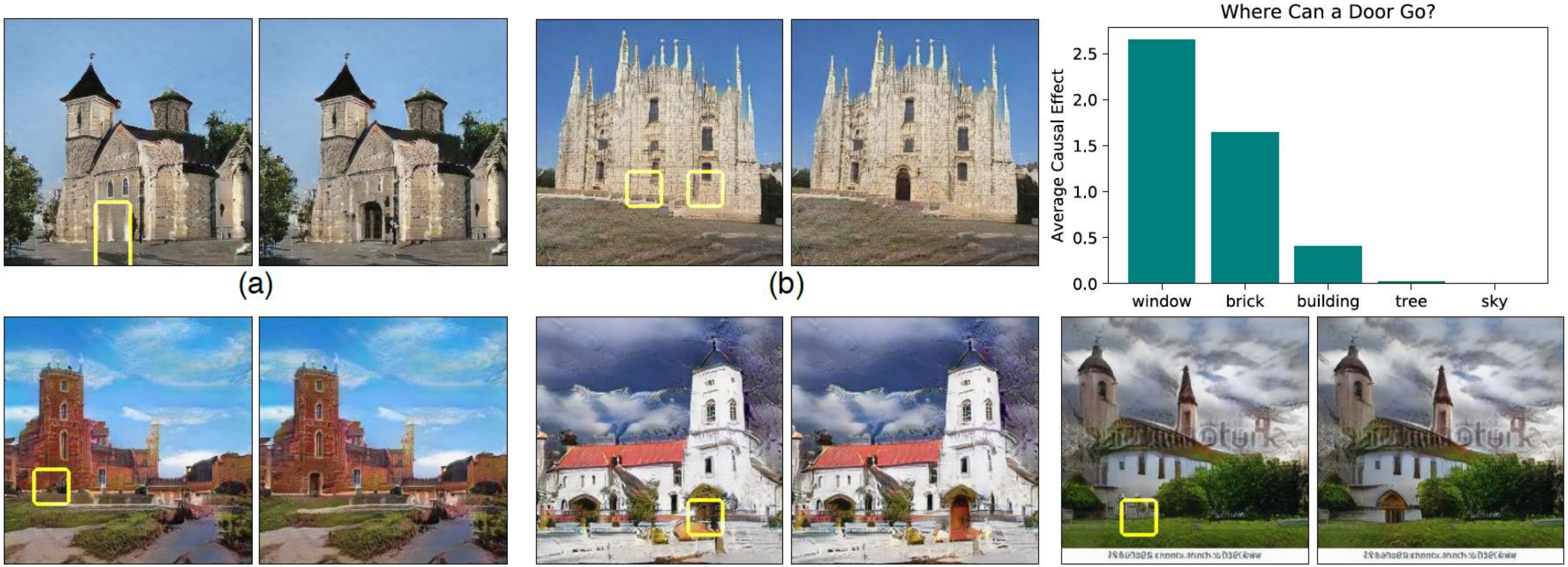
(c) Ablating units removes trees

(d) Activating units adds trees

What it learns

- GAN Dissection

  Effect of Intervention

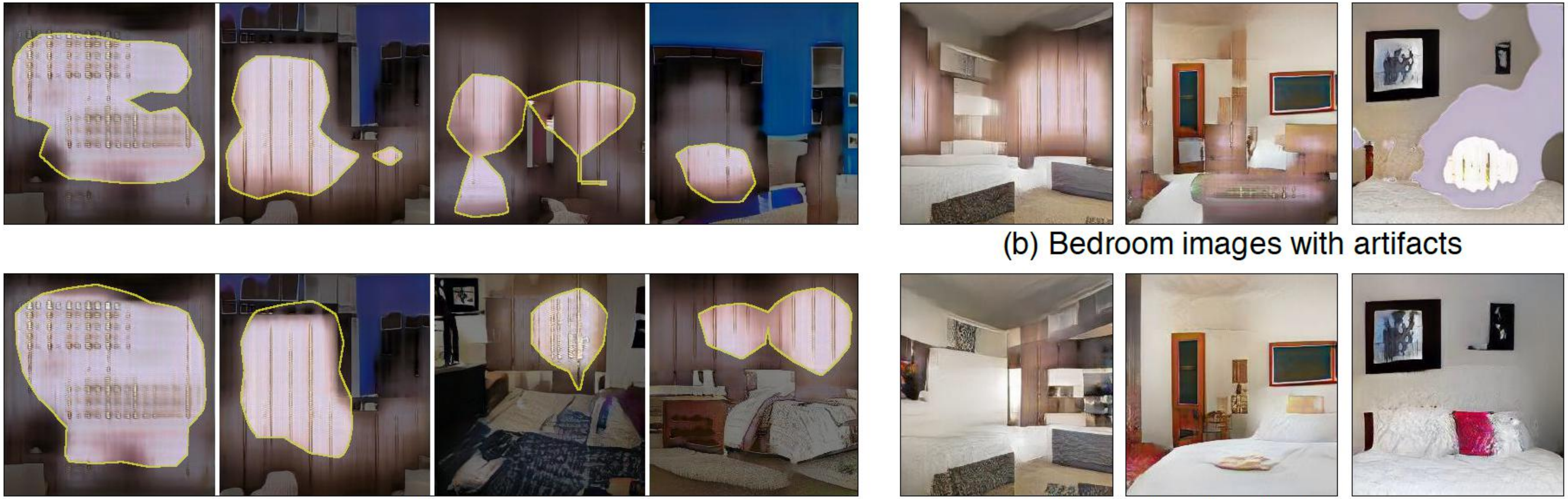


(a) (b) (c) (d) (e)

What it learns

- GAN Dissection

Results



(b) Bedroom images with artifacts

(a) Example artifact-causing units

(c) Ablating “artifact” units improves results

What it learns

- GAN Dissection

https://gandissect.csail.mit.edu

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

What it learns

- Mode Collapse

How do we know what a GAN cannot generate?

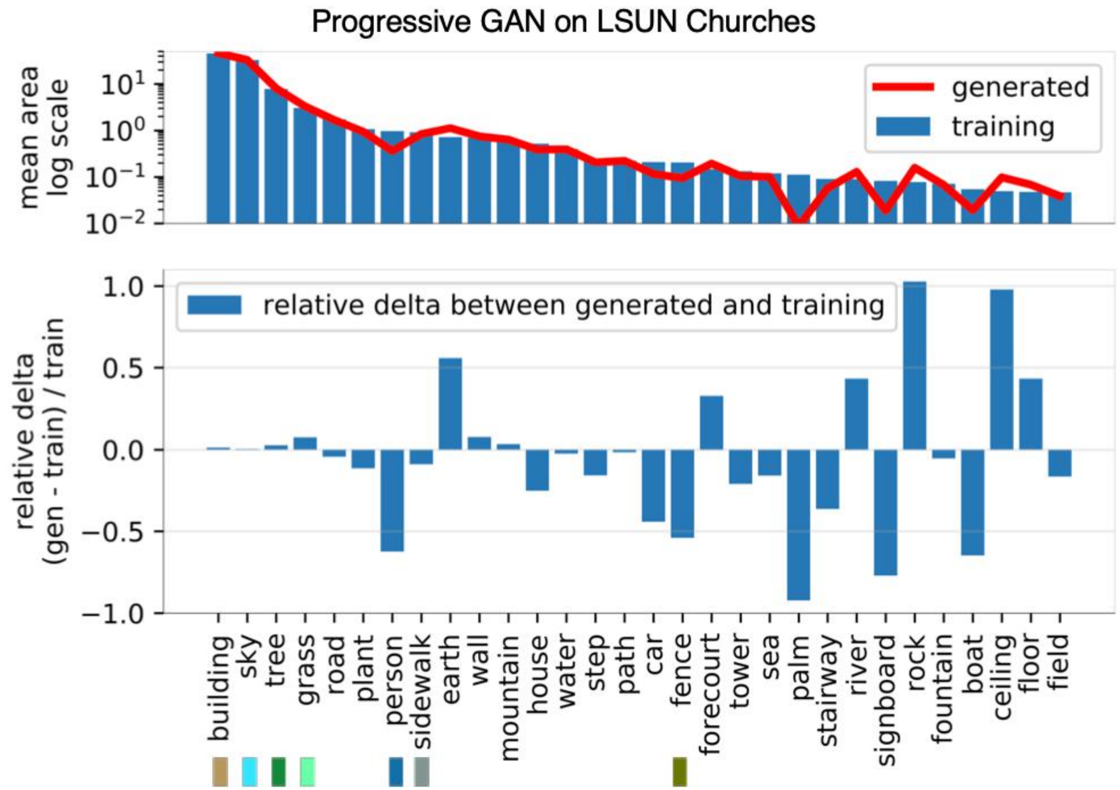
How to visualise the problem of mode collapse ?

Bau, David, et al. "**Seeing what a GAN cannot generate**." Proceedings of the IEEE International Conference on Computer Vision. 2019.

What it learns

- Seeing what a GAN cannot generate

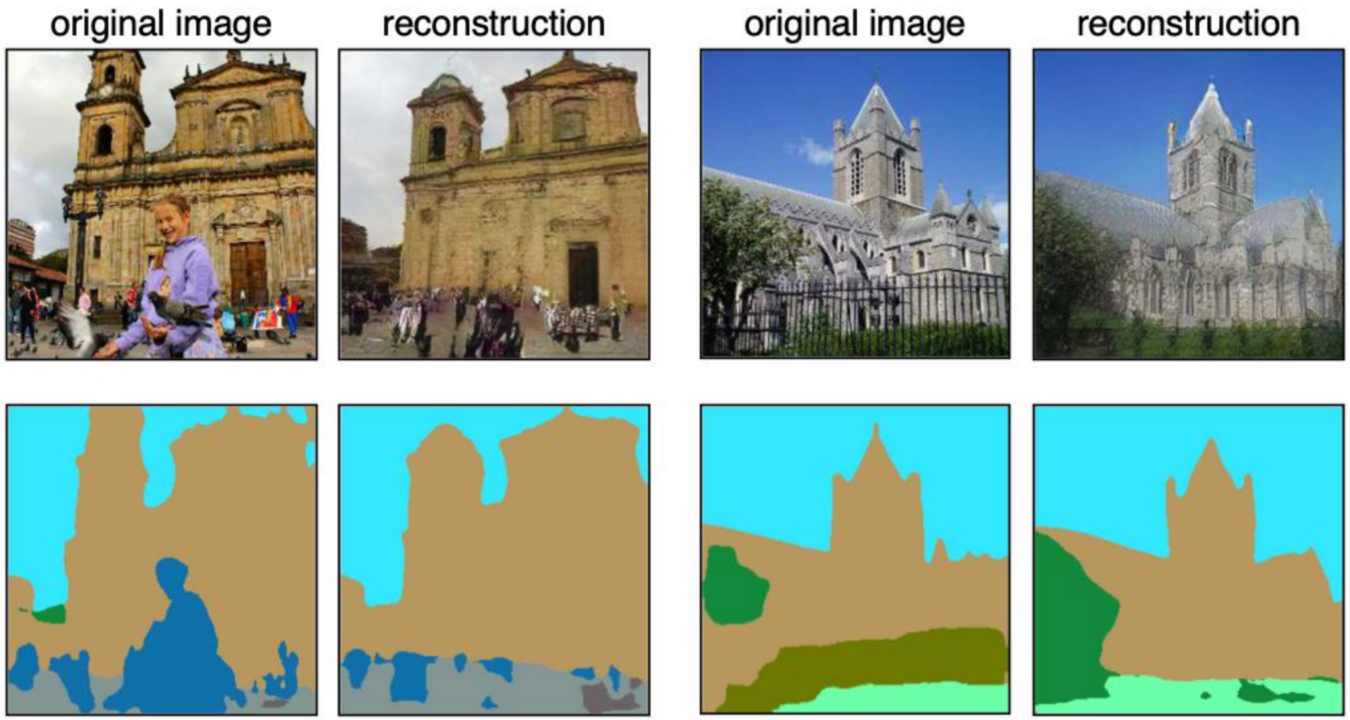
  Generated vs. Training object segmentation statistics



What it learns

- Seeing what a GAN cannot generate

  Generated vs. Training object segmentation statistics

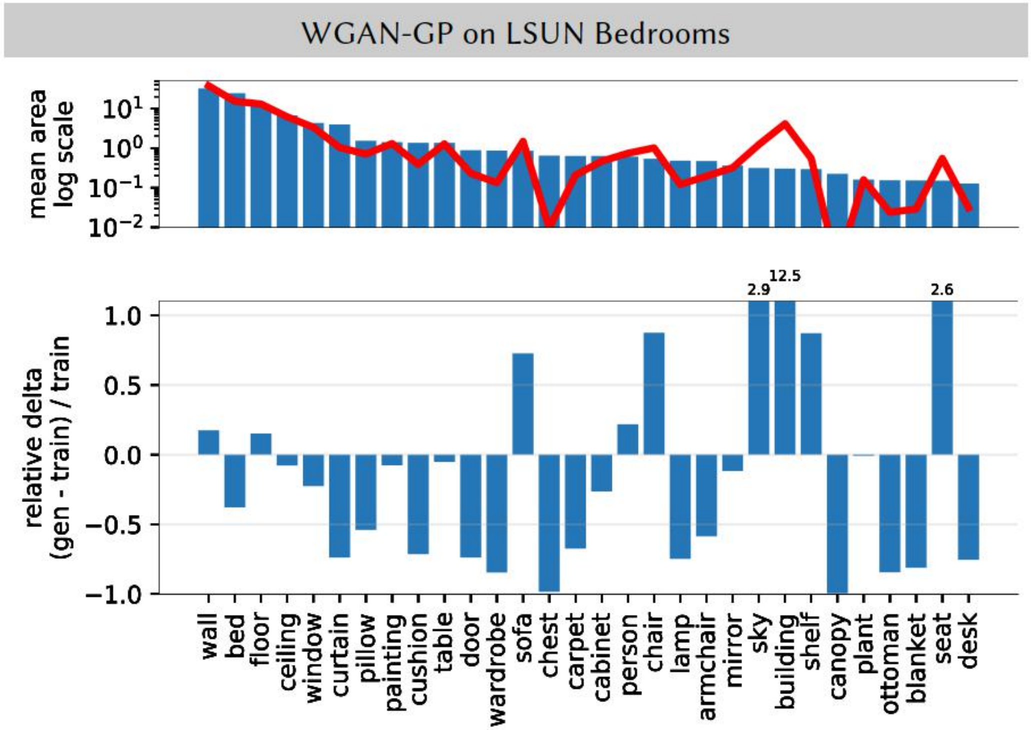


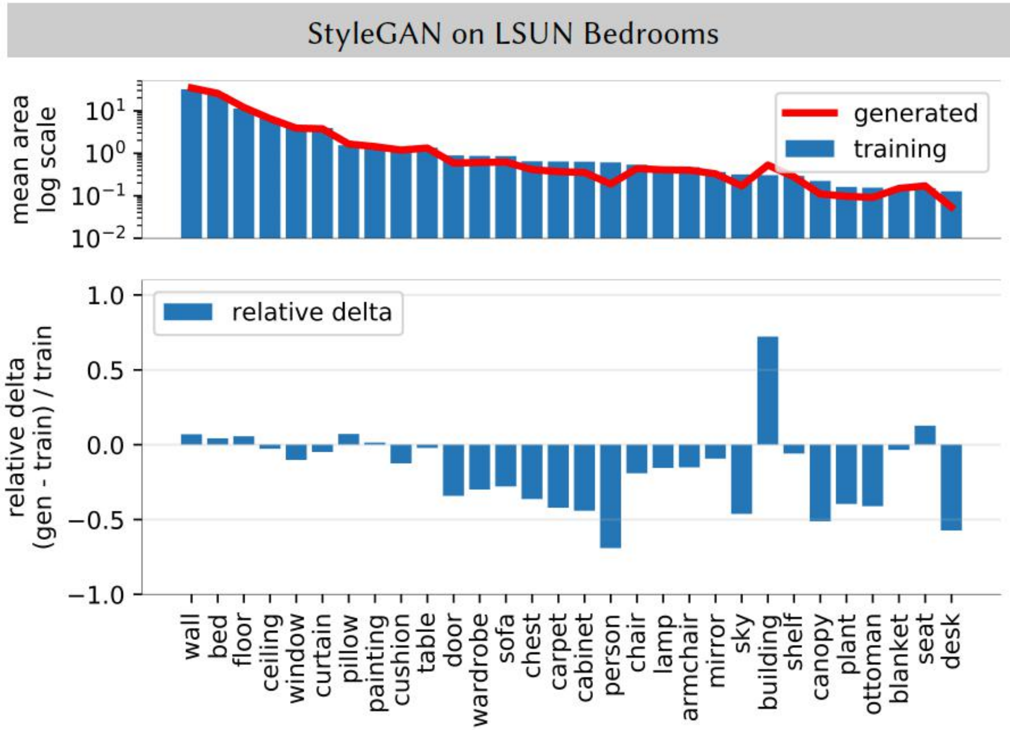
(b) real images vs. reconstructions

What it learns

- Seeing what a GAN cannot generate

  Method: Quantifying **distribution-level** mode collapse

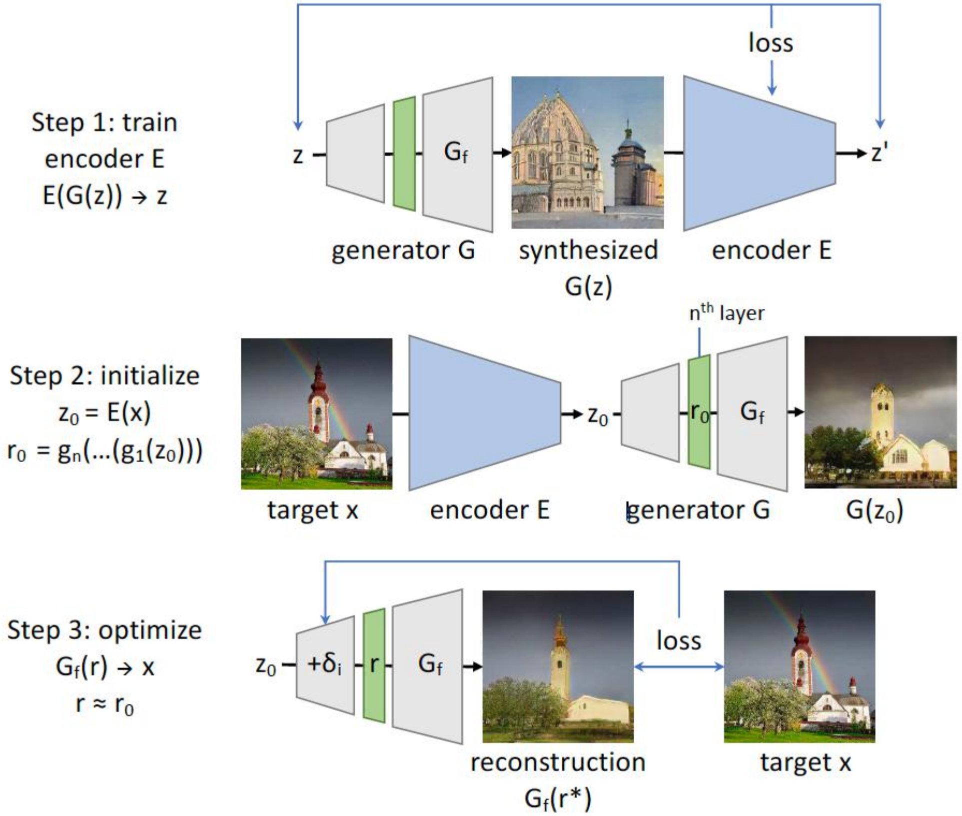




What it learns

- Seeing what a GAN cannot generate

Method: Quantifying **instance-level** mode collapse



$$G = G_f(g_n(\cdots((g_1(\mathbf{z})))))$$

$$\mathcal{L}_{\mathrm{L}} \equiv \mathbb{E}_{\mathbf{z}}[||\mathbf{r}_{i-1} - e(g_i(\mathbf{r}_{i-1}))||_1]$$

$$\mathcal{L}_{\mathrm{R}} \equiv \mathbb{E}_{\mathbf{z}}[||\mathbf{r}_i - g_i(e(\mathbf{r}_i))||_1]$$

$$e_i = \arg\min_e \quad \mathcal{L}_{\mathrm{L}} + \lambda_{\mathrm{R}}\mathcal{L}_{\mathrm{R}},$$

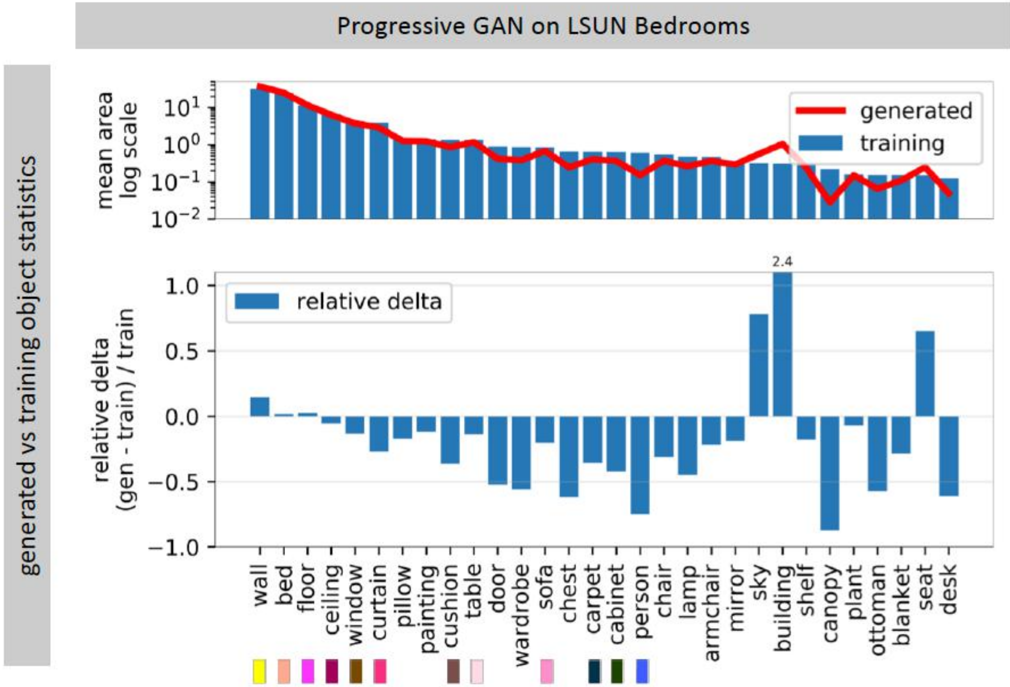
$$E^* = e_1(e_2(\cdots(e_n(e_f(\mathbf{x})))))$$

$$\mathbf{x}' = G_f(\mathbf{r}^*),$$
$$\text{where } \mathbf{r}^* = \arg\min_{\mathbf{r}} \ell(G_f(\mathbf{r}), \mathbf{x})$$

What it learns

- Seeing what a GAN cannot generate

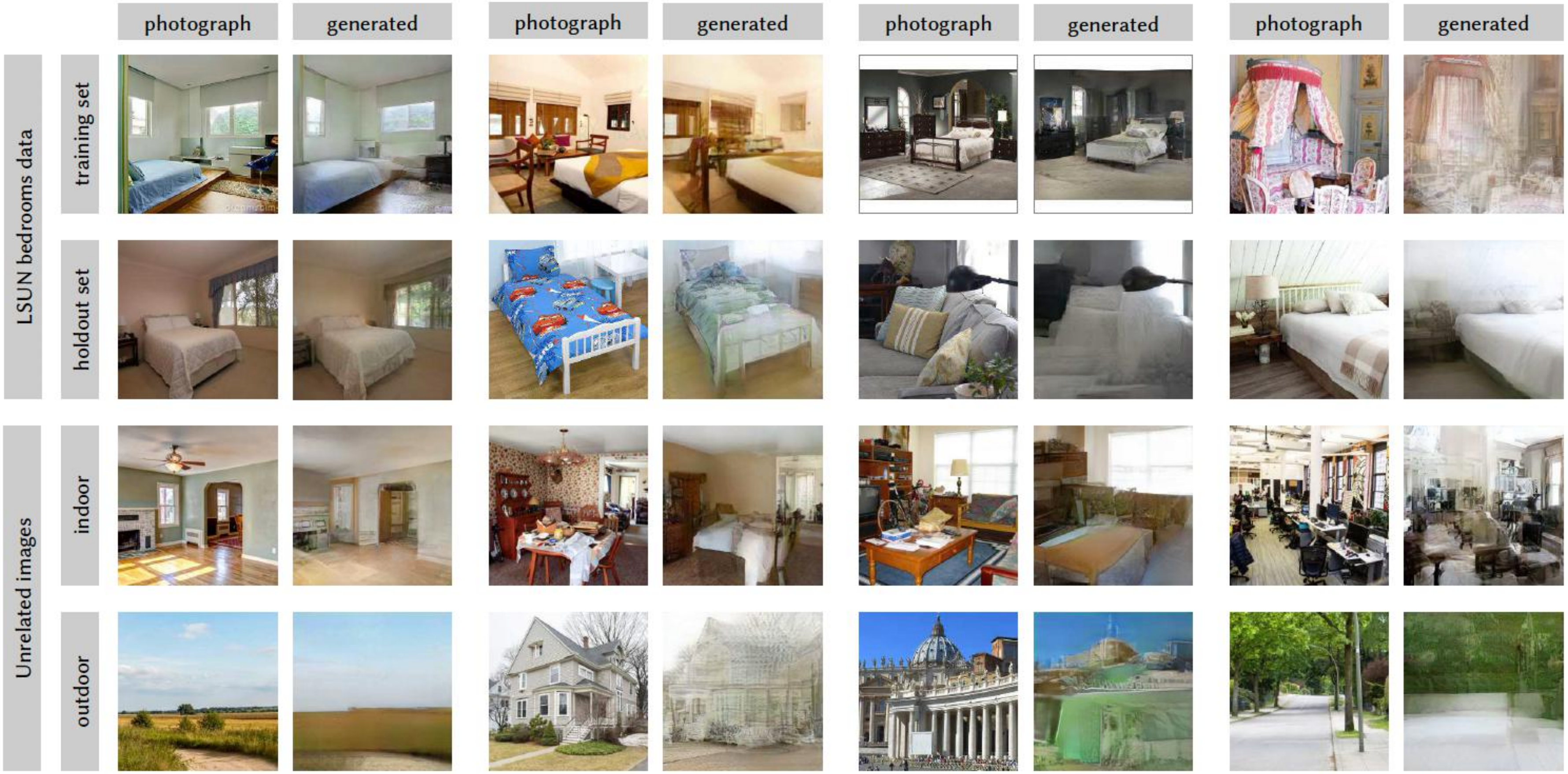
  Results



What it learns

- Seeing what a GAN cannot generate

  Results



What it learns

- Seeing what a GAN cannot generate

Results

Summary

- Internal Distribution Modelling
  - InGAN
  - SinGAN
- What is in the Frequency Domain
  - CNN-generated images
  - Learning in the frequency domain
- What It Learns
  - GAN Dissection
  - Mode Collapse

Thanks